Hypothesis Testing in SEO & Statistical Significance – Whiteboard Friday

A/B testing your SEO changes can bring you a competitive edge and dodge the bullet of negative changes that could lower your traffic. In this episode of Whiteboard Friday, Emily Potter shares not only why A/B testing your changes is important, but how to develop a hypothesis, what goes into collecting and analyzing the data, and thoughts around drawing your conclusions.

Click on the whiteboard image above to open a high resolution version in a new tab!

Video Transcription

Howdy, Moz fans. I'm Emily Potter, and I work at Distilled over in our London office. Today I'm going to talk to you about **hypothesis testing in SEO and statistical significance.**

At Distilled, we use a platform called ODN, which is the Distilled Optimization Delivery Network, to do SEO A/B testing. Now, in that, we use hypothesis testing. You may not be able to deploy ODN, but I still think today that you can learn something valuable from what I'm talking about.

Hypothesis testing

The four main steps of hypothesis testing

So when we're using hypothesis testing, we use four main steps:

1. First, we **formulate a hypothesis.**
2. Then we **collect data** on that hypothesis.
3. We **analyze the data,** and then…
4. We **draw some conclusions** from that at the end.

The most important part of A/B testing is having a strong hypothesis. So up here, I've talked about how to formulate a strong SEO hypothesis.

1. Forming your hypothesis

Three mechanisms to help formulate a hypothesis

Now we need to remember that with SEO we are trying to look to impact three things to increase organic traffic.

1. We're either trying to **improve organic click-through rates.** So that's any change you make that makes yours appearance in the SERPs seem more appealing to your competitors and therefore more people will click your ad.
2. Or you can **improve your organic ranking** so you're moving higher up.
3. Or we could also **rank for more keywords.**

You could also be impacting a mixture of all three of these things. But you just want to make sure that one of these is clearly being targeted or else it's not really an SEO test.

2. Collecting the data

Now next, we collect our data. Again, at Distilled, we use the ODN platform to do this. Now, with the ODN platform, we do A/B testing, and we split pages up into statistically similar buckets.

A/B test with your control and your variant

So once we do that, we take our variant group and we use a mathematical analysis to decide what we think the variant group would have done had we not made that change.

So up here, we have the black line, and that's what that's doing. It's predicting what our model thought the variant group would do if we had not made any change. This dotted line here is when the test began. So you can see after the test there was a separation. This blue line is actually what happened.

Now, because there's a difference between these two lines, we can see a change. If we move down here, we've just plotted the difference between those two lines.

Because the blue line is above the black line, we call this a positive test. Now this green part here is our confidence interval, and this one, as a standard, is a 95% confidence interval. Now we use that because we use statistical testing. So when the green lines are all above the zero line, or all below it for a negative test, we can call this a statistically significant test.

For this one, our best estimate is that this would have increased sessions by 12%, and that roughly turns out to be about 7,000 monthly organic sessions. Now, on either side here, you can see I have written 2.5%. That's to make this all add up to 100, and the reason for that is that you never get a 100% confident result. There's always the opportunity that there's a random chance and you have a false negative or positive. That's why we then say we are 97.5% confident this was positive. That's because we have 95 plus 2.5.

Tests without statistical significance

Now, at Distilled, we've found that there are a lot of circumstances where we have tests that are not statistically significant, but there's pretty strong evidence that they had an uplift. If we move down here, I have an example of that. So this is an example of something that wasn't statistically significant, but we saw a strong uplift.

Now you can see our green line still has an area in it that is negative, and that's saying there's still a chance that, at 95% confidence interval, this was a negative test. Now if we drop down again below, I've done our pink again. So we have 5% on both sides, and we can say here that we're 95% confident there was a positive result. That's because this 5% is always above as well.

3. Analyze the data to test hypothesis

Now the reason we do this is to try and be able to implement changes that we have a strong hypothesis with and be able to get those wins from those instead of just rejecting it completely. Now part of the reason for this is also that we

say we’re doing business and not science.

Here I’ve created a chart of when we would maybe deploy a test that was not statistically significant, and this is based off how strong or weak the hypothesis is and how cheap or expensive the change is.

Strong hypothesis / cheap change

Now over here, in your top right corner, when we have a strong hypothesis and a cheap change, we’d probably deploy that. For example, we had a test like this recently with one of our clients at Distilled, where they added their main keyword to the H1.

This final result looked something like this graph here. It was a strong hypothesis. It wasn’t an expensive change to implement, and we decided to deploy that test because we were pretty confident that that would still be something that would be positive.

Weak hypothesis / cheap change

Now on this other side here, if you have a weak hypothesis but it’s still cheap, then maybe evidence of an uplift is still reason to deploy that. You’d have to communicate with your client.

Strong hypothesis / expensive change

On the expensive change with strong hypothesis point, you’re going to have to weigh out the benefit that you might get from your return on investment if you calculate your expected revenue based off that percentage change that you’re getting there.

Weak hypothesis / cheap change

When it’s a weak hypothesis and expensive change, we would only want to deploy that if it’s statistically significant.

4. Drawing conclusions

Now we need to remember that when we’re doing hypothesis testing, all we’re doing is trying to test the null hypothesis. That does not mean that a null result means that there was no effect at all. All that that means is that we cannot accept or reject the hypothesis. We’re saying that this was too random for us to say whether this is true or not.

Now 95% confidence interval is being able to accept or reject the hypothesis, and we’re saying our data is not noise. When it’s less than 95% confidence, like this one over here, we can’t claim that we learned something the way that we would with a scientific test, but we could still say we have some pretty strong evidence that this would produce a positive effect on these pages.

The advantages of testing

Now when we talk to our clients about this, it’s because we’re aiming really here to give a **competitive advantage over other people in their verticals.** Now the main advantage of testing is to **avoid those negative changes.**

We want to just make sure that changes we’re making are not really plummeting traffic, and we see that a lot. At Distilled, we call that a **dodged bullet.**

Now this is something I hope that you can bring into your work and to be able to use with your clients or with your own website. Hopefully, you can start formulating hypotheses, and even if you can’t deploy something like ODN, you can still use your GA data to try and get a better idea if changes that

you’re making are helping or hurting your traffic. That’s all that I have for you today. Thank you.

Video transcription by Speechpad.com

http://tracking.feedpress.it/link/9375/12943907?utm_medium=feed&utm_source=feedpress.me&utm_campaign=Feed%3A+mozblog